Evaluation for Performance:

Toolkit for Title IV Safe and Drug-Free Schools Programs

2003

James M. O'Neill, Ph.D.
Associate Professor, Psychology Department
Madonna University, Livonia, MI;
O'Neill Consulting, Chelsea, MI

Judith M. Pasquarella, M.P.A.
Manager, Education Section
Office of Drug Control Policy
Michigan Department of Community Health

Henry J. Hastings, J.D., Ed.D.
Acting Director, Michigan Institute for Safe Schools and Communities
College of Education, Michigan State University

This project was supported by the Office of Drug Control Policy (ODCP), Michigan Department of Community Health, with funds from the U.S. Department of Education (UDDOE), Title IV, Safe and Drug-Free Schools and Communities Act. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the ODCP or USDOE. The toolkit may be freely reproduced; however, citation of the source is appreciated.

Suggested reference:

O'Neill, J.M., Pasquarella, J.M., and Hastings, H.J. (2003). Evaluation for performance: Toolkit for Title IV safe and drug-free schools programs. State of Michigan, Department of Community Health, Office of Drug Control Policy.

Chapter 4
Outcome Evaluation:
Assessing Student Performance

What is outcome evaluation?

The purpose of outcome evaluation is to assess short- and long-term changes in participants that result from the *program*. As mentioned earlier, outcome evaluation involves assessment of reaching the destination – the checkered flag – which is your *outcome performance goal/measure*.

This chapter describes the major steps and activities in conducting an outcome evaluation, whereas Chapter 3 covers process evaluation. Please keep in mind that both process and outcome evaluation should be conducted in order to understand fully the value of your program

What are the steps in completing an outcome evaluation?

The focus of this chapter is on evaluating student performance, or the changes in attitudes and/or behavior associated with violence and/or ATOD.

The following table consists of four essential steps in completing your outcome evaluation. The steps involve a process of collecting information about reaching the "destination" of your program, which is expressed as your outcome goal(s) (also known as *performance measure[s]*).

Evaluation Step	Outcome Evaluation Questions
1. Focus on Performance: Use Performance Questions	For each approved outcome performance goal/measure: a. Were the outcome data collected on schedule? If not, provide the reason(s) and a plan to collect the data. b. Were the outcomes in the expected direction? c. Did the outcomes meet or exceed the performance measure? d. Were the outcomes different for various groups (e.g., males vs. females)? e. Were there unintended positive or negative outcomes? f. How clearly were the outcomes attributable to the program?
2. Choose the Best Gauges: Select Indicators, Measures and Sources	a. What outcome *indicator(s)* will be measured to answer the performance questions? b. What *measures* will be used (e.g., survey)? c. Are the measures reliable and valid? d. What information *source(s)* will be used?
3. Check the Gauges - What Do They Say: Collect, Organize and Summarize Information	a. Who will collect the data? When? b. Who will enter/organize the data? When? c. In what format(s) (numbers, words, graphs) will the data be summarized? d. What are the answers to the performance questions in Step 1? e. How and when will the results be reported to stakeholders?
4. Enhance Performance: Make Program Adjustments and Increase Sustainability	a. How will the information be used to enhance the program while preserving *fidelity*? b. How will the information be used to increase sustainability?

The next section provides detailed information about completing each step of the outcome evaluation, followed by a complete example. A checklist is provided at the end of the chapter to use as a roadmap for conducting your own outcome evaluation.

Step 1: Focus on Performance: Use Performance Questions

The standards for outcome evaluation have been converted to the following six questions that each local SDFS coordinator will address in their reports to ODCP. The questions provide a focus which will most likely result in successful outcomes. These questions will be answered for each approved outcome performance goal in your in your grant proposal:

1a. Were the outcomes in the expected direction? The direction of change is a basic yet important indicator, especially when the outcomes did not meet or exceed the performance goal(s), because at least you'll know whether you are moving in the right direction. If the outcomes are in the opposite direction or not as robust as expected, a careful review of the program and process evaluation should occur.

1b. Did the outcomes meet or exceed the performance measure(s) in your grant proposal? This is the primary question of your outcome evaluation, because it relates directly to the approved outcome performance goal(s) of your grant. At a minimum, your answer to this question should include quantitative evidence of change from pre/post results reported as percentages or means and standard deviations. *Qualitative data* are desirable and strongly encouraged, but should not be used in lieu of *quantitative data*.

1c. Were the outcomes different for various groups (e.g., males vs. females)? Not all subgroups may realize similar outcomes, so it is important to report any subgroup differences in outcomes using quantitative data. If available, use qualitative information to further illuminate the observed differences. In addition, any subgroup differences should be considered and monitored as part of program improvements in subsequent years.

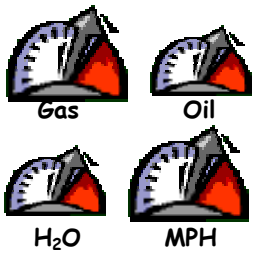
1d. Were there unintended positive or negative outcomes? Not all outcomes can be anticipated, so it's important to identify and report any unintended results. Typically, unintended or negative outcomes emerge and can be understood better when all stakeholders are involved in the interpretation of the results. For example, high prevalence of repeated fighting among a few students may reveal an isolated problem noticed by teachers or parents. Unintended or negative outcomes should lead to program changes that better accommodate students' needs (e.g., the adoption of *selected* or *indicated prevention* programs, in addition to *universal* programs[10]).

1e. How clearly were the outcomes attributable to the program? There are various levels of confidence in attributing student outcomes to programming. Generally speaking, more confidence in the link between programs and outcomes results from implementing scientifically-based programs, because such programs have a track record of effectiveness when implemented with fidelity. A high-performance approach to deduce outcomes from programs is to utilize a *comparison group* or *control group*.[†] These groups have not received the program, but are otherwise similar to the program group (e.g., in age, gender composition and risk status). If the program group changed in the expected direction, but

[†] Essentially, a control group and comparison group serve the same purpose, but a control group is selected through random assignment, whereas a comparison group is chosen through non-random methods. The clearest link between student outcomes and programming is made by using a control group. However, random assignment is a sophisticated process and is not always practical, so check with a professional evaluator for advice and assistance.

the comparison/control group did not, you have strong evidence of successful, program-related outcomes.

Step 2: Choose the Best Gauges: Select Indicators, Measures and Sources



Step 2 is designed to help you select the best and most convenient ***indicators*** (the type of information collected), ***measures*** (the tool used to collect the information) and ***sources*** (the people/places from which to collect the information). Think of the indicators/measures as gauges on a car that give you vital information about your car's status.

2a. What outcome indicators will be measured? The key indicators for an outcome evaluation determined by ODCP include the following: (a) ATOD use, (b) ATOD attitudes, (c) violent behavior and (d) violent attitudes. Changes in these indicators are considered a long-term goal, and you already selected at least one to address in your outcome performance goal, as part of your ODCP grant proposal. However, if you want to replace or add indicators, consider those in the following table. If you decided to change or add indicators, notify your ODCP consultant.

Common Outcome Evaluation Indicators for ATOD and Violence			
ATOD		**Violence**	
Attitudes	**Behaviors**	**Attitudes**	**Behaviors**
☐ Perceived harm/risk from ATOD use ☐ Perceived disapproval in using ATOD ☐ Perceived availability of ATOD ☐ Perceived parental reaction to ATOD use ☐ Pressure to use ATOD	☐ ATOD use in past 30 days ☐ Intensity of ATOD use (e.g., binge drinking) ☐ Frequency of ATOD use ☐ Friends' frequency of ATOD use ☐ Negative experiences (e.g., car crashes, victimization, violence) from own or friends' ATOD use	☐ Perceived harm caused by fighting, bullying or other aggressive behavior ☐ Conditions in which violence is perceived as acceptable ☐ Pressure to engage in violence/aggression ☐ Perception of safety and violence in school	☐Number of fights per student population ☐Number of episodes of harassment, provocations or teasing per student population ☐Number of suspensions for violence, aggression or disrespect per student population ☐Friends' level of violence, aggression or disrespect

Another set of outcome indicators, known as *intermediate outcomes* such as "*risk factors*" and "*protective factors*" can be used to track progress made toward your outcome performance goal(s), much like education benchmarks are used to monitor progress made toward content standards. A list of example risk and protective factors can be found on page 14 of this toolkit, and many survey measures are available online (see Appendix C for further information and URLs).

2b. What measures will be used? You already identified at least one measure as part of an outcome goal in your ODCP grant proposal. If you want to add or change measures, consider the measures on the following page, but review the advantages and disadvantages of each to determine which is appropriate given your expertise and resources.

Type of Measure	Examples	Advantages	Disadvantages
Self-report surveys (Questionnaires=Q; Interviews=I)	Attitudes toward violence; violent behavior; attitudes toward ATOD; ATOD use	• Inexpensive (Q) • Usually high reliability (Q) • Can be anonymous (Q) • Can assess behaviors and attitudes • Can be administered to a large group at one time (Q)	• Validity might be low • Assessment of perceived behavior, not actual behavior • Assess only those who are present • Responses may have high reactivity (e.g., social desirability) • Impersonal (Q) • May need sampling expert
Records	Disciplinary referrals; suspensions; report cards (e.g., code of conduct)	◦ Inexpensive ◦ Usually high validity ◦ Can obtain data for all or a sample of events or participants	◦ May requires extra time for coding and analysis ◦ Information may be incomplete or unclear ◦ Data restricted to what already exists ◦ Access may be limited
Checklists	Teacher and/or parent checklist of student aggression, pro-social behavior	• Direct or indirect assessment of behavior • Usually high validity • Can obtain data for all or a sample of events or participants • Can assess behaviors and attitudes	• May requires extra time for coding and analysis • Requires detailed directions to ensure high inter-rater reliability • Information can be biased by memory, perceptions of rater • May get low response rate from parents, teachers
Observations	Behavior on playground, in classroom, lunchroom or hallway	◦ Direct observation of behavior ◦ Can obtain a lot of detailed information if recorded ◦ High validity	◦ May require extra time for coding and analysis ◦ Requires intensive training to ensure high inter-observer reliability and prevent reactivity

Questionnaires have become a convenient and effective outcome measurement method. In Appendix B you'll find several ready-to-use questionnaires to measure ATOD and/or violent behavior and attitudes. All of these measures are reliable and valid based upon their use in several LEAs. Of course, you should review any measure and test drive it to determine its appropriateness with your targeted population. If the measure needs a little customizing for your population, your ODCP consultant can guide you through that task.

Regardless of the measure used, students have certain rights in an evaluation, including informed consent by the parent/guardian, the right to withdraw from participation in the evaluation at any time, and assurance about the privacy of student information. These safeguards are set forth in the Protection of Pupil Rights Amendment and should be followed carefully.

Helpful Hint
Protecting Student Rights in Program Evaluation

The Protection of Pupil Rights Amendment (PPRA) ensures that the rights of students are protected in evaluations that include methods which reveal information about illegal or anti-social behavior, among other sensitive topics. The federal Department of Education's Family Policy Compliance Office and ODCP can provide detailed guidance and technical assistance regarding implementation of the PPRA - see Appendix C for contact information.

2c. Are the outcome measures reliable and valid?

Reliability and validity are the two primary criteria to assess the quality of your measure. **Reliability** refers to the degree to which a measure is consistent or stable. Using a car analogy, a reliable gas gauge is one that consistently reads empty when the tank is empty. If an outcome measure is unreliable, the "gauge" will not provide a

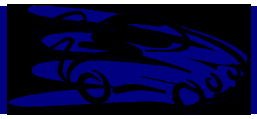
consistently accurate estimate of the attitude or behavior. **Validity**, on the other hand, is the degree to which a measure accurately measures what it's supposed to measure. For example, your gas gauge is supposed to tell you how much gas is in the tank, not how much oil or water. If you are using a measure (e.g., survey) developed by a researcher or evaluator, there's a good chance that it's reliable and valid – just check with the author(s) to make sure and to see if it can be used for the intended population. If you want more information to judge the reliability and validity of your measures, see Appendix B.

You'll Pay Later. . .

It might be tempting to customize an existing measure (e.g., delete items, add items or change item wording) to better accommodate your target population. However, customizing measures can adversely affect their reliability and validity, leading to results that are difficult to interpret. If you think your measure(s) need customizing, consult a professional researcher/evaluator for guidance.

2d. What source(s) of information will be used?

The source(s) of information for outcome indicators can be administrators, teachers, other staff, and students. Whenever possible, obtain information from more than one source. For example, information about violent behavior can be collected from students (e.g., self-report survey) and from teachers (e.g., disciplinary referrals).

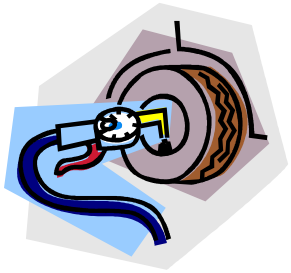


High-Performance Options for Outcome Evaluation

In addition to the basic options for choosing the best measures, choose one or more of the following to enhance your outcome evaluation:

☐ **Use a comparison/control group.** It's the most effective way to attribute outcomes to program efforts. Consult an expert for help.
☐ **Include qualitative data.** Asking a sample of individuals to tell or write about their changes may uncover unanticipated benefits or harms, and show that the original goals were off target.
☐ **Utilize multiple measures.** No one measurement source is perfect; the use of various measures minimizes the weaknesses of any individual measure.
☐ **Collect information at multiple points in time.** It's the best way to capture gradual changes and determine whether changes last.
☐ **Collect information from multiple sources.** The convergence of various sources toward the same results will demonstrate a pattern of real change and enhance the program's credibility with a variety of stakeholders.

Step 3: Check the Gauges – What Do They Say: Collect, Organize and Summarize Information



Checking the gauges includes identifying people who can assist you in collecting and entering/organizing the outcome indicators you identified in Step 2.

3a. Who will collect the data? When? Outcome data could be collected by program facilitators (e.g., teachers) or even those who aren't directly involved in the program (e.g., evaluator), whichever is most feasible.

Regardless of who collects the data, make sure the instructions for data collection are clear and followed by everyone.

Helpful Hints

Outcome Data Collection, Entry and Management

Keep in mind these helpful hints when collecting and entering/managing outcome data:

- **Data collection, entry and management are ongoing processes** that can be time-consuming, depending on the magnitude of your program efforts. Plan enough time and people to complete it in a timely manner.
- **Make data collection routine**. For example, schedule a few minutes on the first and last day of each program series to collect outcome data.
- **Make data collection easy**. Whenever possible, use existing sources (e.g., records) and simple measures (e.g., brief checklists, surveys).
- **Directions for completing measures** should be clear and followed by everyone.
- **Before collecting data, share information with respondents** about the purpose of the data collection, the information requested, the privacy of their responses and their right to withdraw from participating in the evaluation at any time. Keep in mind that the collection of sensitive data (e.g., anti-social and illegal behavior) from students requires adherence to the Protection of Pupil Rights Act. The federal Department of Education and ODCP can provide detailed technical assistance - see Appendix C for contact information.
- **Consider the use of "Scantron" forms** to expedite the data entry process. Contact your district or ISD for more information.
- **Manually collected data** can be entered and managed efficiently in a computer-based "spreadsheet" program.

A timeline for data collection should be completed. There are several outcome data collection designs, each of which provides different information. To choose the most appropriate design, consider the resources you have available and the type of conclusions you want to make about the role of your program in producing the observed outcomes. One the following page is a list of common data collection designs, their description and some comments/recommendations.

Outcome Data Collection Designs (in order of preference suggested by ODCP)		
Name of Design	**Description (example)**	**Comments and Recommendations**
Pre- and Post-Program	Collect data from same persons *prior to* and *after* a program (e.g., collect survey data from the same students before and after their participation in a program)	• Requires use of unique identifiers (e.g., student number) to link data collected at two points in time • Recommended for most programs • Inclusion of a comparison/control group is a "high performance" feature that enhances your ability to link observed outcomes to the program
Pre-, Post-, and Long-Term Post Program	Same as pre- and post-program design, with additional data collected *again at later point(s) in time* (e.g., same as pre- and post-program, plus collect data on same students six months later)	• Same comments as pre- and post-program design, plus requires sustained efforts to conduct long-term follow-up on same students • This is a "high performance" design that allows assessment of *long-term outcomes* • Recommended for LEAs with low-transience populations
Baseline and Follow-Up (Time Series or Time Lag)	Collect year-end data on the same grade, including years before and after a program is instituted (e.g., collect survey data from eighth grade students at the end of each school year for three years before and after a program is instituted)	• Relatively easy to implement because students are not followed longitudinally • Recommended for newly instituted programs, especially where none previously existed, and for programs in which all participants of a unit (e.g., grade, school, district) participate • Inclusion of a comparison/control group is a "high performance" feature that enhances your ability to link observed outcomes to the program • Contact your ODCP consultant before selecting this design
Retrospective Pre- and Post-Program (or Post-Then-Pre Test)	*Following* a program, participants are assessed as they were *prior to* and *after* a program	• Relatively easy to implement because there is only one time of measurement • Recommended for brief programs to avoid over-measurement, or when participants' mistrust or fear of disclosure produces bias in their pre-program responses • Potential bias introduced by relying on respondents' recollection of their pre-program status • Contact your ODCP consultant before selecting this design
Post-Program Only	Data are collected *following* a program; no pre-test or retrospective pre-test is conducted.	• Relatively easy to implement because there is only one time of measurement • Not recommended for outcome assessment because changes are not assessed • Becomes a "high performance" design when used in conjunction with a control group (but not a comparison group) • Contact your ODCP consultant before selecting this design

3b. Who will enter/organize the data? When? A useful way to keep track of data collection and organization/entry is to develop a schedule to identify the following for each indicator: (a) the measure and source used, (b) the data collection person and completion date, and (c) the data organization/entry person and completion date. See page 28 for an example of a completed form which could be used for outcome (and process) indicators (see Appendix C for blank form).

Step 3 also involves summarizing/aggregating information collected on the outcome indicators. Summarizing information can be facilitated by considering the following three questions:

Helpful Hint: Find a Good Mechanic

Make sure you choose a person or a team familiar with various methods to collect and summarize both qualitative and quantitative information. If you need assistance from a professional evaluator, see Chapter 5 for the questions, "How do I find somebody to help evaluate my program?" and "How do I know the evaluator is appropriate for my program?" or contact your funding source(s) for further guidance.

3c. In what format(s) (numbers, words, graphs) will the data be summarized? Data can be categorized in two ways: quantitative (e.g., numbers) and qualitative (e.g., narrative).

For your outcome evaluation, qualitative information should be used to complement – rather than substitute for – quantitative results. Quantitative data are ideal for summarizing large amounts of information, which can be reported using descriptive statistics such as percentages or means and standard deviations. In your report to ODCP and others, descriptive statistics can be presented efficiently in a variety of formats, such as narrative, tables, bar graphs or pie charts.

The use of qualitative information provided from anecdotes, case studies and participant comments is a great way to illuminate quantitative changes – both positive and negative – in student attitudes and/or behaviors.

3d. What are the answers to the outcome performance questions? Using the summary information, answer each outcome performance question provided and developed in Step 1. Remember to be brief, yet concise in your answers and include quantitative evidence (e.g., pre/post statistics, etc.) to support your answers.

3e. How and when will the information be reported to stakeholders? The answers to your outcome performance questions will comprise part of the mid-year and year-end reports submitted to ODCP. The same report format will be used by all local SDFS Coordinators. The report forms, including an example of a completed report, are available in Appendix A and the ODCP web site. In addition, share outcome evaluation results with other stakeholders (e.g., SDFS Advisory Committee/Council, community coalitions, law enforcement, human service agencies) using a presentation format and venue that best suits the target audience.

Step 4: Enhance Performance: Make Program Adjustments and Increase Sustainability

To enhance the performance of your program, it is critical that you:

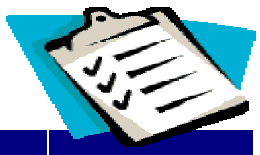
4a. Use the outcome information to improve the program while preserving program fidelity. Use the outcome evaluation results to improve the program. For example, if poor outcomes are found, determine what helped to contribute to those results. Was it poor program planning or training? Poor implementation? Low participant attendance or cooperation? Keep in mind that any program adjustments should not diminish program fidelity. On the other hand, if a careful review of the processes and outcomes shows that the program is ineffective, other scientifically based programs should be instituted.

4b. Use the process information to increase sustainability. Identify ways in which the program can be streamlined to reduce cost or expanded to meet high demand. For example, sharing results about positive outcomes may earn your program a line item in the district budget, especially if the outcomes were valued by and shared with all stakeholders.

What Should I Avoid in Conducting an Outcome Evaluation?

The following are common problems in outcome evaluation, accompanied by potential solutions:

Problem (example)	You'll Pay Later Because. . . 	Solution(s)
Use of district-wide survey data as outcome results for a small group of participants (e.g., use district-wide results from the *Western Michigan AOD Survey* for a program implemented in only one school/classroom).	The results for the program participants will be masked by that of non-participants.	• Collect data using district-wide survey for program participants only. • Implement program in a sufficient number of schools/classrooms to make the district-wide survey results applicable.
The sample size for data collection is too small and/or did not include all program sites (e.g., use sample of 10 students in one school as outcome results for a program of 100 participants in three schools).	The sample will not be representative of the larger program group(s), which will lead to biased outcome conclusions.	• Include staff in preparation of data collection plan, and provide incentives for data collection. • Screen completed surveys for completeness before including them in the analysis.
A lot of missing pre/post survey matches (e.g., most students completed the pretest survey, but many did not complete the posttest).	The sample will not be representative of pre-post changes in the larger program group(s), which will lead to biased outcome conclusions.	• Provide incentive for completing survey. • Collect data during "captive audience" times. • Compare pre-test scores between participants who do and do not have a posttest. If the pre-test scores are similar, you can cautiously conclude that the pre-post results would be similar for those who didn't complete the posttest.
Suspected bias on self-report surveys (e.g., under- or over-reporting of ATOD use.)	The results will not reflect the true behaviors/attitudes of the respondents, which will lead to biased outcome conclusions.	• Increase trust between staff and students. • Use surveys with clearly worded items that measure recent behavior. • Collect data from secondary sources (e.g., office referrals). • For post-only or post-then-pre, make it anonymous. • For pre-post survey, use code numbers and emphasize confidentiality.
Using process evaluation results as evidence of outcome effectiveness (e.g., use high participation rates or widespread popularity of the program to show that the program "works").	Process evaluation results, though important, cannot provide evidence of successful student outcomes (behaviors/attitudes), which is the ultimate goal of the program.	• Refer to Chapter 2 for the distinction between process evaluation and outcome evaluation. • Include outcome indicators to demonstrate program effectiveness. • Utilize process evaluation results to demonstrate how the process led to the observed outcomes.



Outcome Evaluation Checklist

Outcome Evaluation Step And Suggested Timeline		Person(s) responsible	Date Complete	Notes:
Step 1: Focus on Performance Complete in year prior to program implementation.				
Jan. – Apr.	☐ Meet with Advisory Council/Committee to review Performance Questions and outcome evaluation steps.	Advisory Council	/ /	
Apr. – May	☐ Develop additional performance questions, if needed.		/ /	
Step 2: Choose the Best Gauges Complete in year prior to program implementation.				
Apr. – May	☐ Select the best *indicators* (the types of information) to be collected		/ /	
Apr. – May	☐ Select the best *measures* (the tools) used to collect the information.		/ /	
Apr. – May	☐ Select the most appropriate *sources* (the people/places) from which to collect the information.		/ /	
Apr. – May	☐ For high-performance outcome evaluation, choose one or more options, p. 37.		/ /	
Ongoing	☐ Avoid/remedy common problems in outcome evaluation (see p. 41).		/ /	
Step 3: Check the Gauges Complete during program implementation.				
Apr. – May	☐ Identify who will collect the data for the various indicators, and by when.		/ /	
Pre-Post program	☐ Collect/Organize the data in a routine, timely manner. Refer to helpful hints on p. 38.		/ /	
May – Aug.	☐ Summarize the data based upon the performance questions to be answered.		/ /	
Aug.– Sep.	☐Use the summarized results to answer each performance question concisely and completely (using the ODCP report forms).		/ /	
Jan. & Sept.	Report your answers to the performance questions to ODCP.		/ /	
Jan. & Sept.	☐ Share results with your local Advisory Council/Committee and other stakeholders.		/ /	
Step 4: Enhance Performance Complete during and after program implementation.				
Oct. – Dec.	☐ Use outcome results to make adjustments to the program while preserving fidelity.		/ /	
Oct. – Dec.	☐ Use outcome results to secure additional support and/or resources		/ /	

Notes: